R E P O R T R E S U M E S

ED 011 071 24

THE RELIABILITY OF SPECIAL TESTS IN MEASURING PERSONALITY.

BY- ONG, JIN
CENTRAL WASHINGTON STATE COLL., ELLENSBURG

REPORT NUMBER BR-5-8174 PUB DATE JUN 66
REPORT NUMBER CRP-S-358
EDRS PRICE MF-$0.09 HC-$1.00 25P.

DESCRIPTORS- *BEHAVIOR RATING SCALES, *PERSONALITY ASSESSMENT, *TEST VALIDITY, *TEST RESULTS, *TEST SELECTION, ELLENSBURG

A FOLLOWUP STUDY WAS REPORTED THAT ENLARGED THE SCOPE OF THE AUTHOR'S PREVIOUS STUDY OF THE OPPOSITE-FORM APPROACH USED BY STUDENTS IN TEST AND MEASUREMENT COURSES. THE STUDY HAD THREE PURPOSES--(1) TO INVESTIGATE THE RELIABILITIES OF OPPOSITE-FORM INVENTORIES, (2) TO CROSS VALIDATE OPPOSITE-FORM INVENTORIES, AND (3) TO STUDY THE PATTERNING OF EXTERNAL VALIDITY OF THE OPPOSITE-FORM APPROACH. THE FOLLOWUP STUDY ADDED INFORMATION ON THE TEST-RETEST RELIABILITY OF THE OPPOSITE-FORM APPROACH AND SHOWED ITS VALUE AND PRACTICAL SIGNIFICANCE TO PERSONALITY INVENTORIES IN AN INVESTIGATION OF THE PATTERNING VALIDITY USING A PRACTICAL EDUCATION CRITERION (COURSE GRADE). SUBJECTS CONSISTED OF 96 MALE AND 95 FEMALES, A TOTAL OF 191 FRESHMEN, SOPHOMORES, AND JUNIOR COLLEGE STUDENTS ENROLLED IN PSYCHOLOGY CLASSES. FIVE INVENTORIES WERE GIVEN IN FOUR SETTINGS, EACH ABOUT 2 DAYS APART. THE IMPLICATION OF THE RESULTS WAS THAT THE CORRELATION OR PREDICTION OF AN EDUCATION CRITERION IS JUST AS EFFECTIVE IN USING THE OPPOSITE FORMS OF AN INVENTORY AS IN THE ORIGINAL. (AL)

ED011071

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

THE RELIABILITY OF SPECIAL TESTS IN MEASURING PERSONALITY

Cooperative Research Project No. 5-8174
(S-358)

Jin Ong

Central Washington State College
Ellensburg, Washington

1966

This research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

TABLE OF CONTENTS

Page

List of Tables.....iii

Problem.....1

Objectives.....1

Related Research.....1

Procedure.....6

Analyses of the data and Findings.....7

Conclusions and Implications.....9

References.....11

Appendix: Tables.....13

List of Tables

1. Comparison of Ong's Odd-Even Correlated r's between Opposite-Forms A and B.....13

2. Comparison of Ong's Kuder-Richardson Correlated r's between Opposite-Forms A and B.....14

3. Various Reliability Coefficients of Opposite-Forms of Five Inventories.....15

4. Comparison of Test-Retest r's between Two Sub-Groups..16

5. Listing of the Opposite-Form r's of Ong's Earlier and Present Investigations.....17

6. Comparison of Opposite-Form r's between Ong's Earlier and Present Groups.....18

7. Validity Coefficients of the Opposite-Form Inventories with Course Grade.....19

8. Comparison of Correlated r's between Opposite-Forms A vs. B, A' vs. B', All of these Forms Correlated With Course Grade.....20

Problem

The present study had three general purposes: first, to investigate the reliabilities of opposite-form inventories, second, to cross validate the opposite-form inventories, and third, to study the patterning of the external validity of the opposite-form approach.

Objectives

1. To compare the internal consistency reliabilities of Opposite Forms A and B: a) the odd-even correlations, and b) the Kuder-Richardson correlations.
2. To determine the test-retest reliabilities of Forms A and B.
3. To cross validate the test-retest reliabilities of Forms A and B, using two sub-groups.
4. To cross validate or to compare the opposite-form reliabilities obtained in the present study with those of an earlier study.
5. To determine the validity coefficients of the opposite-form inventories with an educational criterion.
6. To compare the correlated r's between Forms A vs. B, A' vs. B' all of these forms correlated with an education criterion.

Related Research

The development of Opposite forms may be seen from the following. In 1923, Cady made a selection of questions from the Woodworth material, with considerable modification of phrasing. In order to give the questions an additional effectiveness, another form was prepared in which every question in the first form was reversed in the other. Cady noted a lack

of clearness and directness in many questions. He admitted that it was very difficult to change a statement from the natural direct way of expression into a question that said the same thing yet required an opposite reply. The result was a number of expressions which were practically double negatives. At any rate, the various reliabilities of his questionnaire ranged form .25 to .42. He concluded that it is open to question whether a single test containing items in direct and reversed forms would be as valuable as the same number of carefully selected questions divided into two forms and given at different periods. In 1925 Laird described the Colgate Mental Hygiene Test, and Hoitsma (1925) gave some statistical results from its use. This inventory was an extension of the Woodworth (1918) material with fifty additional items dealing with introversion and extroversion. Here the main interest centers on the introversion and extroversion additions, since these parts were dealing with some kind of opposite forms of items. At any rate, Hoitsma did report the r's of -.22 to -.45 between introvert and extrovert scores. At about the same time Symonds (1925) constructed a social attitudes questionnaire to measure liberalism. In order to test the reliability of the questionnaire, a second form was prepared containing the same questions so expressed that the answers would be the opposites to the answers appropriate for the items in the original questionnaire. The correlation of the two forms was .67. After 1925 the practice of constructing reversed forms was not employed until the appearance of Adorno's book on The Authoritarian Personality in 1950. The main concern here

was the F-Scale in The Authoritarian Personality. A group of writers attempted to construct a reversed form of the F-Scale to suit the purposes of their own investigation. Thus once again the practice of constructing reversed forms was in vogue.

Among this group of writers, Chapman and Campbell (1957) employed two sets of ten and sixteen items in their original form, and about the same number in the reversed form. The correlations between the scores on original and reversed items on two occasions were only .17 and -.01. Chapman and Campbell concluded that the achievement of a successful reversal would be a very difficult task. Other writers also had ideas of reversed form construction for this inventory. Jackson and Messick (1957) constructed the reversed F-Scale by rewriting each item so that the content would appear to reflect a viewpoint opposed to the original, while retaining a similar style of expression. Later, Jackson, Messick, and Solley (1957) reported a correlation of .35 between their original and reversed forms. Bass (1955) constructed the G-scale on the basis that each new statement was opposite in meaning--or as opposite as the author could make it--to an original statement in the G-scale. Bass found that the r between F- and G-scales was -.20. Leavitt, Hax, and Roche (1955) selected half of the F-scale items and reworded them into negative forms so that authoritarianism would now be indicated by disagreement rather than agreement with the items, and found that the r's between the reversed and unreversed halves with five small samples were between -.42 and .45, except one r of -.66 with an N of 6.

These studies on the low r's between the original and the reversed F-scales were also cited and discussed by Christie, Havel and Seidenberg (1958), who attempted to construct a reversed F-scale based on the criteria of logical and psychological opposition, and the avoidance of extremity of wording. The correlations between the scores from the fifteen original and the fifteen reversed items were between .14 and .45. At about the same time Spivack (1956) developed an instrument to measure certain aspects of self-acceptance and self-rejection. Upon examination of Spivack's inventory, it was interesting to discover that the inventory contained both a positive and a negative form--a set of opposite forms. Although this inventory consisted of the items to measure self-acceptance and self-rejection, Spivack apparently had not touched upon the idea of opposite forms since nowhere was this idea mentioned, nor were correlations given between the self-acceptance and self-rejection items.

In general, investigators, with the exception of Symonds, have reported a low relationship or an absence of relationship between the original and the reversed forms as constructed by each investigator according to his particular concept. At this point the following questions seem pertinent: Why was there such a low relationship, or an absence of relationship, as found by these researchers? Why was Symonds' finding an exception? Very recently, Ong (1965), in his book called, The Opposite-Form Procedure in Inventory Construction and Research, has attempted to explain the rationale behind the low relationship between the original and reversed forms found by these investigators and the reason behind Symonds' apparent success in obtaining a moderate relationship. Ong also has investigated certain aspects

of the effectiveness and flexibility of oppositely-stated items for use in constructing personality inventories. He has refined and developed certain concepts of oppositeness, discussed some related research on some aspects of the concepts of opposite forms, analyzed the items of opposite forms of an inventory, illustrated the practical construction of the opposite forms of seven inventories, and carried out a series of experimental studies using the constructed opposite forms of these inventories. A general conclusion was that opposite forms of the inventory type containing items with opposite directional idea could be constructed with moderately high reliability.

There has been a fashion for the modern writers to construct opposite forms not for the main purpose of research in opposite form construction, but for the purpose of investigating an area called "acquiescence" or "response style" in personality inventories. Rorer (1963) has reviewed at length studies in this area. It seems that the inventory used most frequently in the construction of such forms is the MMPI, or its derivatives. The most recent examples are those of Rorer and Goldberg (1965 a, b), who constructed the reversed MMPI items in an attempt to argue for the non-existence of acquiescence. It seems that their correlations might have been influenced by the fact that they found items in their reversed MMPI which were quite ineffective to attain reversal. Also they have not yet performed the retest on their reversed MMPI.

Earlier, Ong (1965) has demonstrated certain effective procedures in opposite or reversed form construction. But there are certain special areas remained to be investigated, for example, the retest using the opposite forms, the comparison of the internal consistency reliability between the opposite forms, the cross validation of the opposite forms, and the correlation or prediction of an educational criterion using opposite forms.

Procedure

The opposite-forms A and B of five inventories constructed by Ong (1965)--Test I, Carter's (1958) Mechanics of Study Procedure (good items); Test II, Allport's (1928) Ascendant-Submissive Behavior; Test III, Gough's (1953) Honor Point Scale; Test IV, Edward's (1957) Social Desirability Scale; and Test V, Spivack's (1956) Self-acceptance and Self-rejection (good items)--were given to a general psychology class and three other classes in development, learning and evaluation. Subjects consisted of 96 males and 95 females, a total of 191 freshmen, sophomores, and juniors enrolled in the author's classes in one quarter. The five inventories were given in four sittings, each about two days apart, that is, test Form A, test Form B, retest Form A', and retest Form B. The order of giving these four forms was as follows:

Class	I		II		III		IV	
	T.H.	B.H.	T.H.	B.H.	T.H.	B.H.	T.H.	B.H.
1st round	A	B	A	B	A	B	A	B
2nd round	B	A	A'	B'	B	A	A'	B'
3rd round	A'	B'	B	A	A'	B'	B	A
4th round	B'	A'	B'	A'	B'	A'	B'	A'

For example, in the first round, the top half of the students in each of four classes (top half vs. bottom half alphabetically) took Form A first while the bottom half took Form B first; in the second round, the procedure was reversed for two classes of the four, that is, the top half took Form B, while the bottom half took Form A. The remaining two classes took the retest; that is, top half took Form A', while bottom half took Form B', etc. The inventories were scored as discrepancies with the key. The course grade for each student was calculated in terms of T-scores, using the average

T-scores of about ten weekly examinations. The odd-even r's and the Kuder-Richardson r's from the earlier study (Ong, 1965) were used to calculate the significance of the correlated r's between the opposite-forms A and B. The various reliability coefficients of the Forms A and B, A' and B' of the five inventories were calculated for the present sample. Also the correlations between the Forms A and B, A' and B', all with course grade, were calculated. Then various comparisons were made for different correlations.

Analyses of the Data and Findings

Table I presents comparison of Ong's (1965) odd-even correlated r's between the Opposite Forms A and B. Out of the five tests there exists a significant difference in only one test, that is, Test V. Table 2 again gives comparisons of Ong's Kuder-Richardson correlated r's between the Opposite Forms A and B. None of the differences is significant. Consequently, with one significant difference out of ten comparisons in internal consitency reliabilities between Opposite Forms A and B, one may have some moderately high degree of confidence that these two forms are comparable.

Table 3 lists the various reliability coefficients of the Opposite Forms of the five inventories of the present sample. All the r's are significant and way above .01 level. In general, the test-retest r's seem to be a little higher than the opposite-form r's. These results conform to the usual observation that test-retest r's are higher than the parallel-form r's. Table 4 presents the cross validation or comparison of the test-retest reliabilities of Forms A and B, using the two

largest sub-groups (freshmen vs. juniors). None of these differences between the test-retest r's, which were converted to z's, was significant. Thus these results show that these Opposite-Forms are quite comparable from one group to the other, as far as the college population is concerned.

Table 5 lists the opposite-form r's of Ong's earlier and the present investigation. Since the retest was not done in the earlier study, the retest r's are listed only for the present study. Also the r's projected to 100 items are listed. Table 6 presents a cross validation or comparison of the opposite-form reliabilities obtained in the present college sample with those of an earlier adult-extension sample. Again none of these differences is significant. Thus these results show that the Opposite-Forms are quite comparable, at least from undergraduate college sample to the adult-extension sample.

Table 7 lists the validity coefficients of the opposite-form inventories with course grade. The r's of four out of five tests are significant at least at the .05 level. The minus r's conform with the intent of the test constructor since the inventories were scored as discrepancies with the key. The results show that Tests I, III, IV and V correlated or predicted the criterion, course grade, about equally well, and at least better than Test II. Yet, the correlation or prediction is not too much above significance, or only several percent better than chance. Table 8 presents the comparisons of the correlated r's between Opposite Forms A vs. B, A' vs. B', all of these forms correlated with course grade. Since in Test II, only one r, that is, .16, is barely significant out of four,

comparison is not made for this test. Again, none of these differences in the remaining four tests is significant. Consequently, the opposite-forms are comparable, as far as the validity coefficients using an outside educational criterion is concerned. The implication of these results seems to be that the correlation or prediction of an educational criterion, course grade, is just as effective in using the opposite-forms of an inventory as is the original.

Conclusions and Implications

Within the limits of these samples, the following conclusions and implications seem warranted:

1. With one significant difference out of ten comparisons in internal consistency reliabilities between the opposite-forms, the comparability of these forms was considered to be moderately high.
2. All the various reliability coefficients, including the test-retest r's, were highly significant.
3. There was no significant difference in the cross validation or comparison of test-retest reliabilities of the opposite forms, using a freshman and junior sample.
4. There was no significant difference in the cross validation or comparison of opposite-form reliabilities obtained in the present college sample with those of an earlier adult-extension sample.
5. The validity coefficients of the opposite-form inventories with an educational criterion, course grade, were significant for four out of five inventories. These four tests correlated or predicted course

grade about equally well, but only about several percent better than chance.

6. There was no significant difference between the correlated r's of the opposite forms correlated with course grade.
7. As a general conclusion, these opposite forms using several measures of comparison were highly comparable.
8. The implication of these results seems to be that the correlation or prediction of an educational criterion, course grade, is just as effective in using the opposite-forms of an inventory as is the original.

References

Adorno, T. W., et al. The authoritanian personality. New York: Harper and Bros., 1950.

Allport, G. W. A test for ascendance-submission. J. Abn. & Soc. Psychol., 1928, 23, 118-136.

Bass, B. M. Authoritarianism or acquiescence. J. Abn. & Soc. Psychol., 1955, 51, 616-623.

Cady, V. M. Estimation of juvenile incorrigibility. J. Delinq., Mongr. No. 2, 1923.

Carter, H. D. The mechanics of study procedure. Calif. J. Educ. Res., 1958, 9, 8-13.

Chapman, L. J., and Campbell, D. T. Response set in the F scale. J. Abn. & Soc. Psychol., 1957, 54, 129-132.

Christie, R. H., Havel, J. and Seidenberg, B. Is the F-scale irreversible? J. Abn. & Soc. Psychol., 1958, 56, 143-159.

Edwards, A. L. The social desirability variable in personality assessment and research. New Hork: The Dryden Press, 1957.

Gough, H. G. The construction of a personality scale to predict scholastic achievement. J. Applied Psychol., 1953, 37, 361-366.

Hoitsma, R. K. The reliability and relationships of the Colgate Mental Hygiene Test. J. Applied Psychol., 1925, 9, 293-303.

Jackson, D. N., and Messick, S. J. A note on "ethnocentrism" and acquiescent response sets. J. Abn. & Soc. Psychol., 1957, 54, 132-134.

Jackson, D. N., Messick, S. J., and Solley, C. M. How "rigid" is the "authoritarian?" J. Abn. & Soc. Psychol., 1957, 54, 137-140.

Laird, D. Detecting abnormal behavior. J. Abn. Psychol., 1925, 20 128-141.

Leavitt, H. J., Hax, H., and Roche, J. H. Authoritarianism and agreement with things authoritative. J. Psychol., 1955, 40, 215-221.

Ong, J. The opposite-form procedure in inventory construction and research. New York: Vantage Press, 1965.

Rorer, L. G. The great response style myth. Oregon Res. Inst. Monogr., Vol. 3, No. 6, 1963.

Rorer, L. G. and Goldberg, L. R. Acquiescence in the MMPI? Educ. & Psychol. Meas't., 1965, 23, 801-817.

Rorer, L. G. and Goldberg, L. R. Acquiescence and the vanishing variance component. J. Applied Psychol., 1965, 49, 422-430.

Spivack, S. S. A study of a method of appraising self-acceptance and self-rejection. J. Genetic Psychol., 1956, 33, 183-202.

Symonds, P. M. A social attitudes questionnaire. J. Educ. Psychol., 1926, 16, 316-322.

Woodworth, R. S. Woolworth's Personal Data Sheet. Chicago: C. H. Stoelting Co., 1918.

TABLE 1

Comparison of Ong's Odd-Even Correlated r's

between Opposite Forms A and B

(N = 61 adult extension students)

Test	Form	Odd-Even r	r^2	D_r	r_{AB}	r^2_{AB}	t	P
I	A	.04	.0016	.06	.67	.4489	.57	N
	B	.10	.01					
II	A	.60	.36	.06	.83	.6889	1.05	N
	B	.66	.4356					
III	A	.49	.2401	.06	.72	.5184	.672	N
	B	.43	.1849					
IV	A	.42	.1764	.09	.77	.5929	1.18	N
	B	.51	.2601					
V	A	.52	.2704	.20	.67	.4489	2.70	.01
	B	.72	.5184					

TABLE 2

Comparison of Ong's Kuder-Richardson Correlated

r's between Opposite Forms A and B

(N = 61 Adult extension students)

Test	Form	K-R r	r^2	D_r	r_{AB}	r^2_{AB}	t	P
I	A	.19	.0361	.04	.67	.4489	.388	N
	B	.23	.0529					
II	A	.73	.5329	.02	.83	.6889	.398	N
	B	.71	.5041					
III	A	.41	.1681	.12	.72	.5184	1.344	N
	B	.29	.0841					
IV	A	.59	.3481	.04	.77	.5929	.592	N
	B	.63	.3969					
V	A	.72	.5184	.07	.67	.4489	1.001	N
	B	.65	.4225					

TABLE 3

The Various Reliability Coefficients of

Opposite-Forms of Five Inventories

(Total N = 191)

	Test				
Forms	I	II	III	IV	V
AA'	.75	.85	.80	.84	.84
BB'	.57	.89	.70	.81	.76
AB	.53	.75	.73	.74	.72
A'B'	.70	.83	.68	.78	.73
AB'	.66	.79	.69	.73	.68
BA'	.61	.79	.72	.76	.73

TABLE 4

Comparison of Test-retest r's between two Sub-groups.

Test		Forms	r	z	D_z	t	P
I	Freshman N=75	AA'	.74	.95	.08	.43	N
	Junior N=52		.70	.87			
	Freshman	BB'	.64	.76	0	0	N
	Junior		.64	.76			
II	Freshman	AA'	.82	1.16	.14	.76	N
	Junior		.77	1.02			
	Freshman	BB'	.86	1.29	.13	.70	N
	Junior		.89	1.42			
III	Freshman	AA'	.86	1.29	.29	1.57	N
	Junior		.76	1.00			
	Freshman	BB'	.74	.95	.10	.54	N
	Junior		.78	1.05			
IV	Freshman	AA'	.78	1.05	.02	.11	N
	Junior		.79	1.07			
	Freshman	BB'	.80	1.10	.03	.16	N
	Junior		.81	1.13			
V	Freshman	AA'	.83	1.19	.19	1.03	N
	Junior		.76	1.00			
	Freshman	BB'	.75	.97	.24	1.30	N
	Junior		.62	.73			

$\sigma_{d_z} = .185$

TABLE 5

Listing of the Opposite-Form r's of Ong's

Earlier and Present Investigations

		r		r projected to 100 items	
	Test	AB	A'B'	AB	A'B'
Earlier Investigation					
N = 61 Adult university Extension students	I	.67		.91	
	II	.83		.96	
	III	.72		.93	
	IV	.77		.94	
	V	.67		.91	
Present Investigation					
N = 191 Freshmen, Sophomores, and Juniors	I	.53	.70	.85	.92
	II	.75	.83	.94	.96
	III	.73	.68	.93	.91
	IV	.74	.78	.93	.95
	V	.72	.73	.93	.93

TABLE 6

Comparison of Opposite Form r's between

Ong's Earlier and Present Groups

Test		Opposite-Form	r	z	D_z	t	P
I	Earlier N=61	AB	.67	.81			
					.22	1.47	N
	Present N=191	AB	.53	.59			
		A'B'	.70	.87	.06	.40	N
II	Earlier	AB	.83	1.19			
					.22	1.47	N
	Present	AB	.75	.97			
		A'B'	.83	1.19	0	0	N
III	Earlier	AB	.72	.91			
					.02	.13	N
	Present	AB	.73	.93			
		A'B'	.68	.83	.08	.53	N
IV	Earlier	AB	.77	1.02			
					.07	.47	N
	Present	AB	.74	.95			
		A'B'	.78	1.05	.03	.20	N
V	Earlier	AB	.67	.81			
					.10	.67	N
	Present	AB	.72	.91			
		A'B'	.73	.93	.12	.80	N

$\sigma_{d_z} = .15$

TABLE 7

Validity Coefficients of the Opposite-Form

Inventories with Course Grade

(N= 191)

	TEST				
Form	I	II	III	IV	V
A	-.24*	.16*	-.25*	-.22*	-.21*
A'	-.24*	.11	-.24*	-.21*	-.25*
B	-.19*	-.02	-.21*	-.17*	-.28*
B'	-.18*	.06	-.22*	-.18*	-.16*

Minimum sig. r's: with N = 191

.15 at .05

.19 at .01

The inventories were scored as discrepancies with the key

TABLE 8

Comparison of Correlated r's between Opposite-Form A vs. B, A' vs. B', All of these Forms Correlated with Course Grades (N=191)

Test	Form	r	r^2	D_r	$r_{AB \text{ or } A'B'}$	$\left(r_{AB \text{ or } A'B'}\right)^2$	t	P
I	A	-.24	.0576	.05	.53	.2809	.711	N
	B	-.19	.0361					
	A'	-.24	.0576	.06	.70	.4900	1.06	N
	B'	-.18	.0324					
II	A	.16						
	B	-.02						
	A'	.11						
	B'	.06						
III	A	-.25	.0625	.04	.73	.5329	.751	N
	B	-.21	.0441					
	A'	-.24	.0576	.02	.68	.4624	.345	N
	B'	-.22	.0484					
IV	A	-.22	.0484	.05	.74	.5476	.948	N
	B	-.17	.0289					
	A'	-.21	.0441	.03	.78	.6084	.617	N
	B'	-.18	.0324					
V	A	-.21	.0441	.07	.72	.5184	1.30	N
	B	-.28	.0784					
	A'	-.25	.0625	.09	.73	.5329	1.69	N
	B'	-.16	.0256					

* Since only one r, that is, .16, is barely significant and the remaining three r's not significant, comparison is not made for this test.